Comparison of sequences generated by a hidden Markov model

George Kerchev (Université du Luxembourg)

Time **Thursday, Oct 10, at 16:00**
Place **Campus Kirchberg, room MNO 5A (5th floor)**

The length LC_n of the longest common subsequences of two strings $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$ is a way to measure the similarity between X and Y. We study the asymptotic behavior of LC_n when the two strings are generated by a hidden Markov model $(Z, (X, Y))$ and we build upon asymptotic results for LC_n obtained for sequences of i.i.d. random variables. Under some standard assumptions regarding the model we first prove convergence results with rates for $\mathbb{E}[\mathrm{LC}_n]$. Then, versions of concentration inequalities for the transversal fluctuations of LC_n are obtained. Finally, we outline a proof for a central limit theorem by building upon previous work and adapting a Stein's method estimate.